
Early Results on Blue Gene at SDSC

Wayne Pfeiffer
February 24, 2005



SAN DIEGO SUPERCOMPUTER CENTER

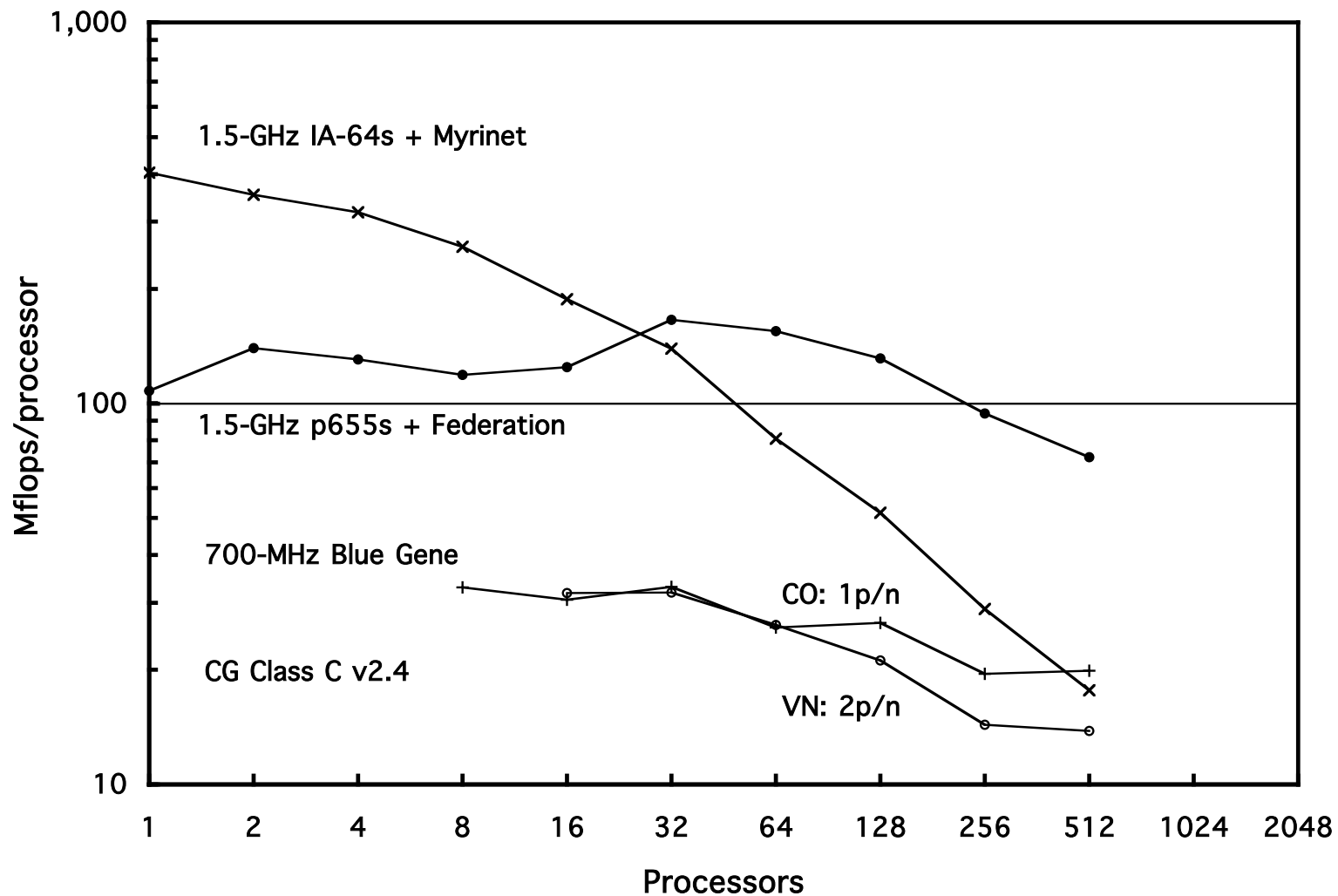
at the UNIVERSITY OF CALIFORNIA, SAN DIEGO



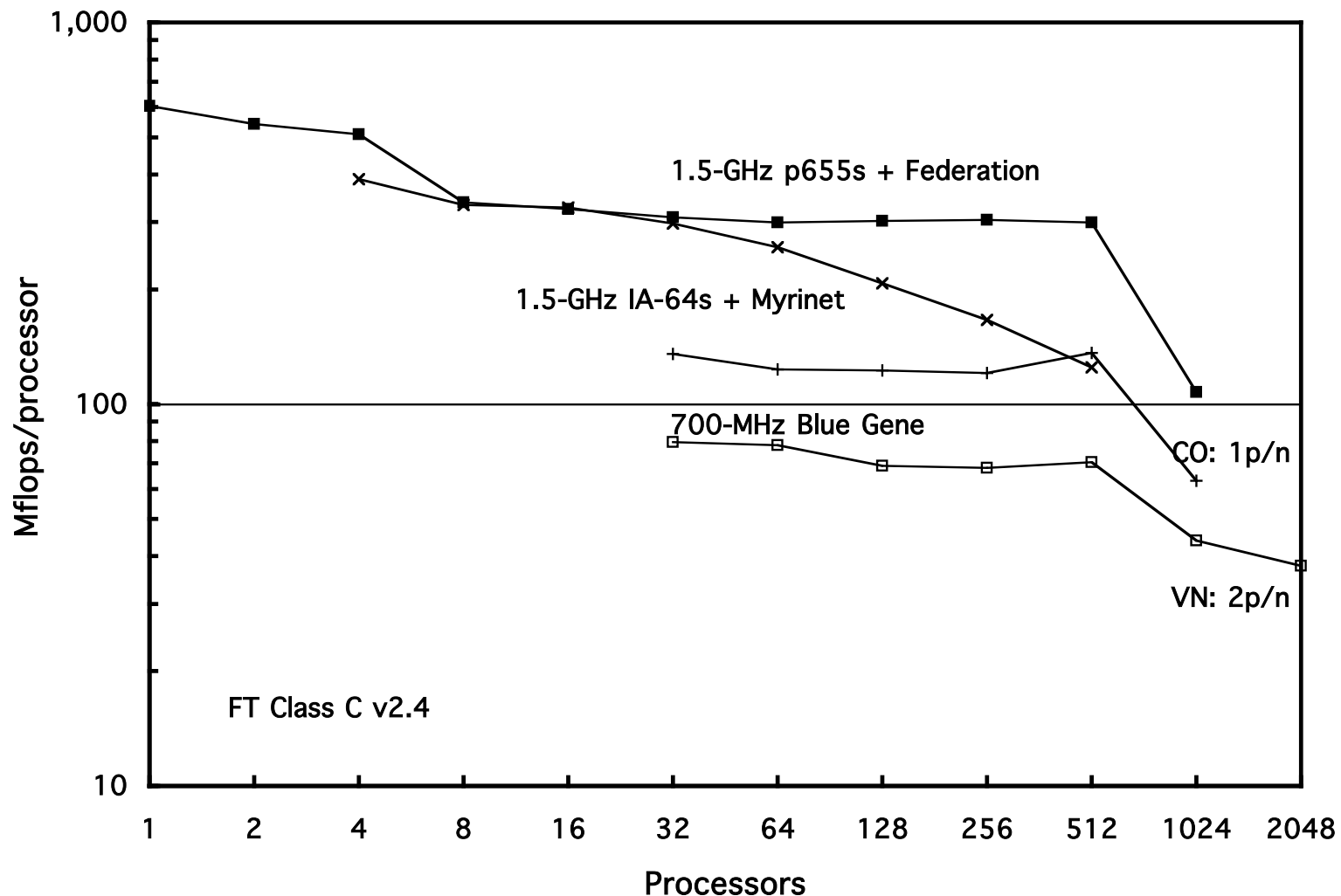
Strong scaling results were obtained on Blue Gene for four NPB 2.4 C kernels & NAMD 2.5

- **NPB kernels were compiled with no changes**
 - Options were -O3 -qarch=440d for CG, MG, & LU
 - Options were -O5 -qnopia -qarch=440d for FT
- **NAMD required numerous changes to compile**
 - Changes were made by student from UIUC visiting IBM
 - Options were -O -qarch=440 (i.e., no optimization to avoid NaNs)
- **mpirun (with Driver 521 for NPBs & 480 for NAMD) used**
 - -partition to specify block (still generally necessary for VN mode)
 - -mode CO & -mode VN for comparison
 - -env "BGLMPI_EAGER=128" for NAMD runs
- **Results were compared with previous ones on**
 - DataStar with 1.5-GHz p655s + Federation
 - TeraGrid with 1.5-GHz IA-64s + Myrinet

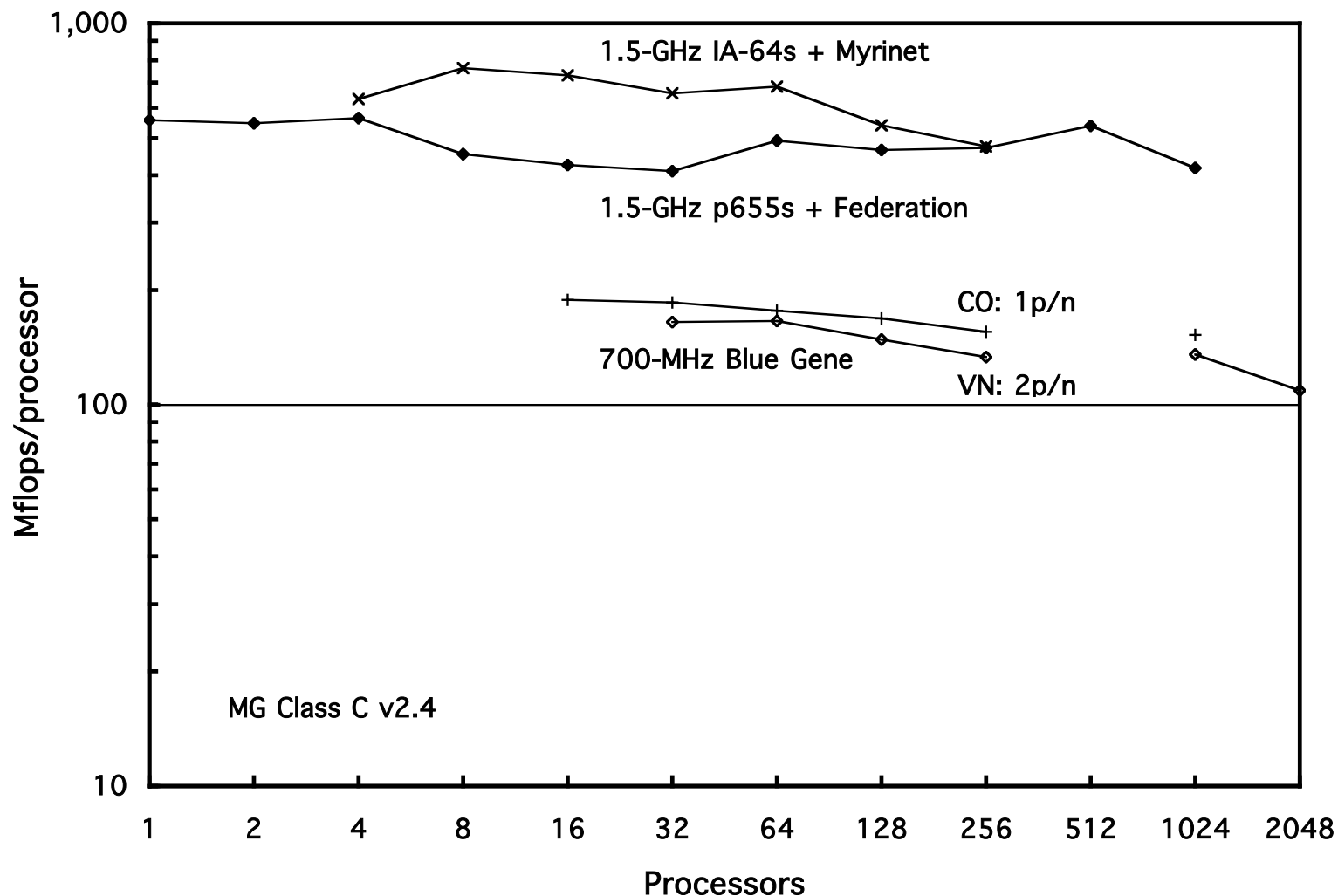
*CG scales to 128p or so on BG & p655s, but badly on IA-64s;
VN mode is much worse than CO mode (per p) for $\geq 128p$; BG is slow*



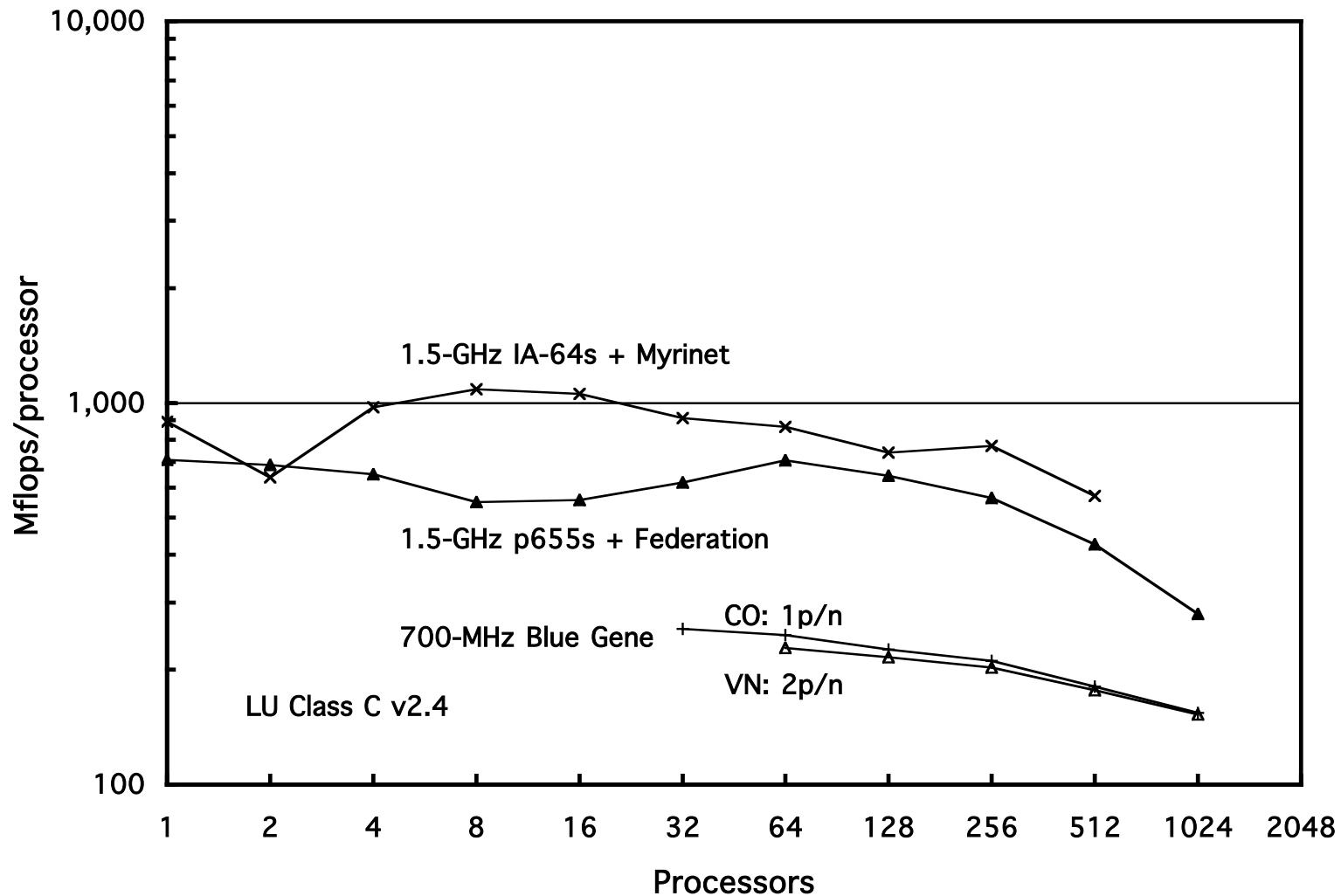
***FT scales well to 512p on p655s & BG, but not on IA-64s;
VN mode is much worse than CO mode (per p) for all p***



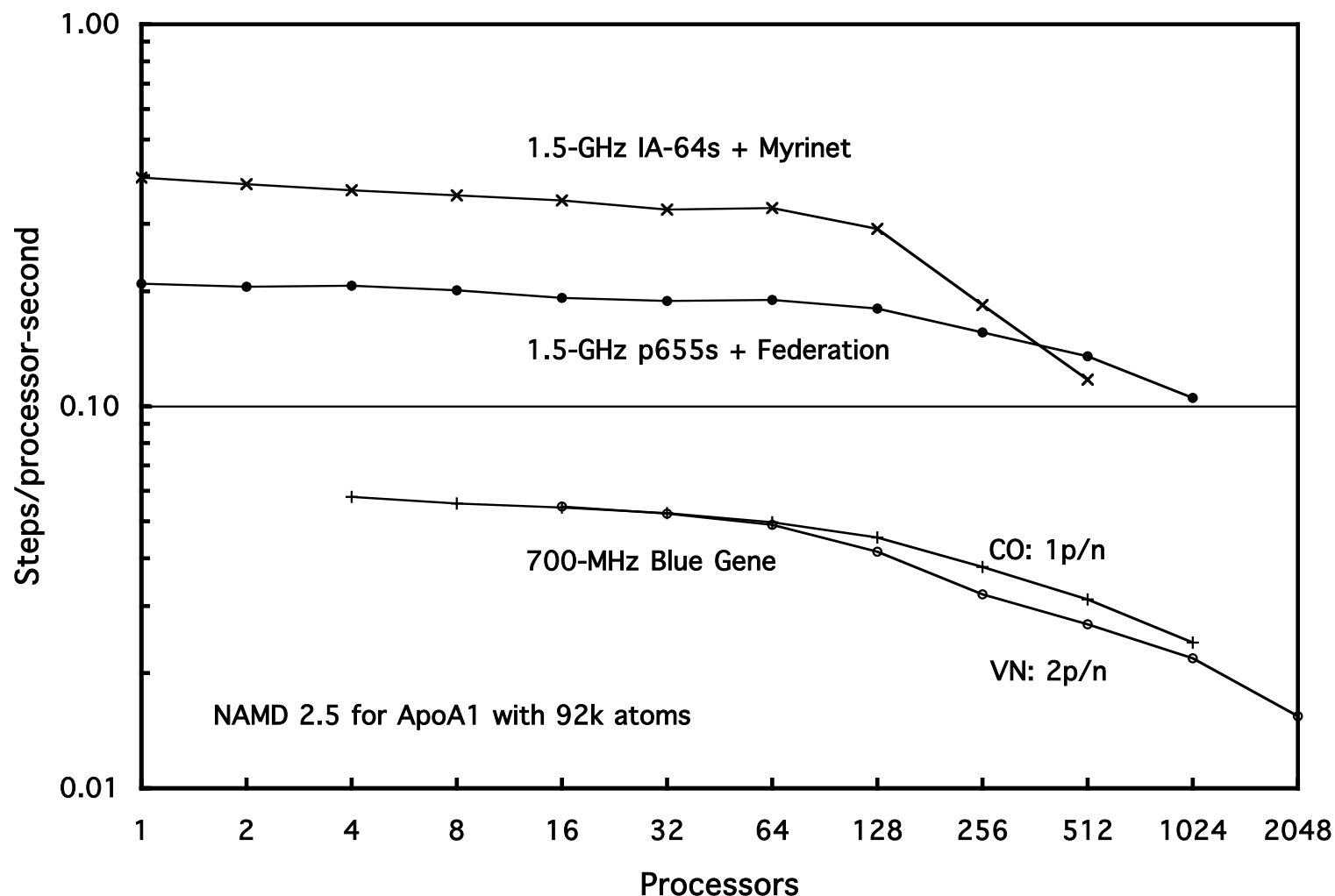
*MG scales well on all machines; 512p case fails on BG & IA-64s;
VN mode is moderately worse than CO mode (per p) for all p*



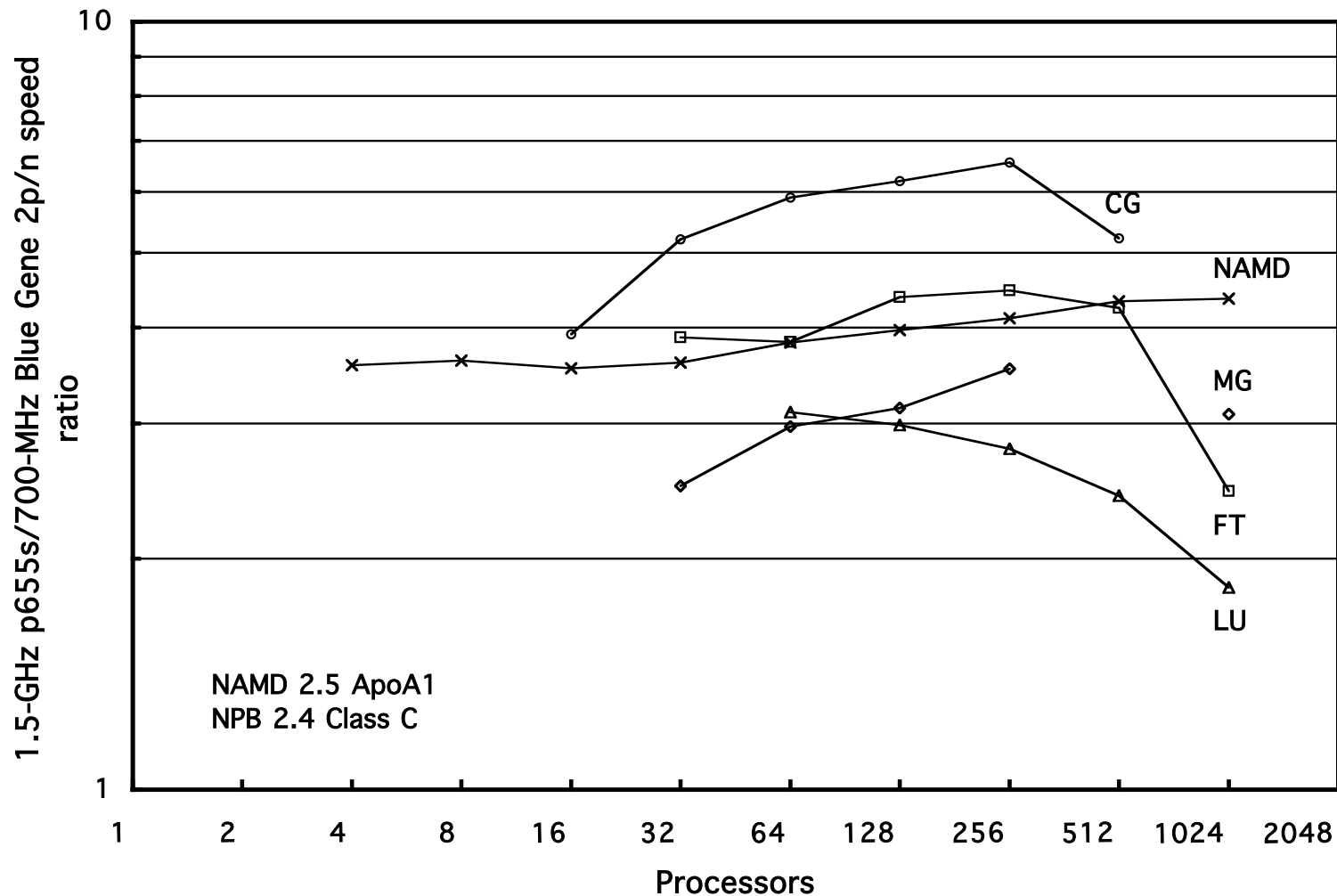
*LU scales well to 1024p on BG & 512p on p655s & IA-64s;
VN mode is about the same as CO mode (per p) for all p*



*For NAMD, so far, BG doesn't scale as well as p655s;
VN mode is only a little worse than CO mode (per p)*



Speed of p655s compared to Blue Gene with VN mode is highest for CG & lowest for LU



Changes in mpirun settings were investigated for NPB 2.4 C kernels

- **No explicit -partition generally faster for $\leq 256p$ (only in CO mode now)**
 - Faster with no -partition for all kernels on 64p, 128p, & 256p (presumably because more links are available)
 - by 1.12x to 1.14x for FT & 1.05x to 1.09x for CG
 - About the same for other processor counts, except
 - Slower with no -partition by 0.96x for CG on 32p (?)
- **-connect TORUS generally faster than -connect MESH (only if -partition not specified & only in CO mode now)**
 - Faster with TORUS by 1.11x to 1.16x for FT on 64p to 1024p
 - Slightly faster with TORUS for CG & MG on most processor counts
 - No difference for LU
 - Slower with TORUS by 0.94x for CG on 32p (?)

Changes in driver & compiler flags were also investigated for NPB 2.4 C kernels

- **Driver 521 vs 480 speed the same except**
 - Faster with 521 by 1.15x for FT on half rack (512p) in CO mode
 - Faster with 521 by 1.09x for FT & 1.05x for MG on half rack (1024p) in VN mode
- **-O5 -qnoipa vs -O3 speed the same except for FT**
 - Faster with -O5 -qnoipa by 1.03x for FT on 32p to 512p in CO mode
 - Faster with -O5 -qnoipa by 1.05x for FT on 32 & 64p in VN mode
 - Slower with -O5 -qnoipa by 0.95x for FT on 2048p in VN mode

Other codes are being ported &/or tested

- **CPMD: quantum molecular dynamics**
 - Running executable from IBM
 - Need full rack to have enough memory for problem of interest
- **DOT: protein docking**
 - Tracking down NaNs during execution
- **DNSMSP: 3-D turbulence**
 - Need to change FFT algorithm from 1-D to 2-D decomposition to use less memory & scale better
- **ENZO: 3-D cosmology**
 - Need to eliminate some scaling bottlenecks to run problems of interest
 - Need fast, parallel I/O
- **SPECFEM3D: 3-D seismic wave propagation**
 - Compiles with -O5 -qarch=440d
 - Have trouble fitting in memory of single rack
 - Need fast, parallel I/O

System software priorities for SDSC

- **mpirun with anomalies and hangs fixed**
- **LoadLeveler for batch job submission**
- **GPFS for fast, parallel I/O**
- **HPC Toolkit for performance analysis**
- **TotalView debugger**